Rainbow Spaghetti, Metro Maps, and Object-Centric Process Mining

Wil van der Aalst

Merriam-Webster defines a **metaphor** as a "figure of speech in which a word or phrase literally denoting one kind of object or idea is used in place of another to suggest a likeness or analogy between them". Metaphors can be used to create vivid imagery, exaggerate a characteristic or action, or express a complex idea. Examples are "Life is a journey, so enjoy each step along the way" and "Time is like a river, constantly flowing, unyielding and relentless". In **process mining**, we often use metaphors to clarify concepts. Some examples:

- A **discovered process model** is often described as an "**X-ray** of the organization" showing the process that would otherwise not be visible. Like an X-ray, the process model shows the problem before deciding on surgery or some other form of treatment.
- Process mining starts from **event data**, and traditionally, events are grouped into **traces**, i.e., sequences of activity executions. Such traces can be seen as "**desire lines**". A desire line is an informal path created over time by people choosing the most direct or preferred route, often diverging from designated pathways. This metaphor nicely explains that the actual behaviors of people

and organizations often differ from the intended, designed, or expected behaviors.

- Twenty years ago, I already spoke about "**Spaghetti processes**" to describe the **complexity and lack of structure** in process models discovered using process mining.
- In my **CAiSE 2009** (International Conference on Advanced Information Systems Engineering) **keynote**, I spoke about "TomTom for Business Process Management (TomTom4BPM)". In the keynote, I compared processes to traffic flows, process models to geographic maps, and information systems to navigation. Since then, I have been using Google Maps and Metro Maps as metaphors for process models. Note that **Google Maps** can be overlayed with traffic information (congestion, roadblocks, etc.), just like we can overlay process models with performance and compliance indicators (e.g., average waiting time and frequency).

Recently, we witnessed an uptake of **Object-Centric Process Mining (OCPM).** OCPM fundamentally changes the process mining discipline, and its impact should not be underestimated. Therefore, I exploit metaphors like Rainbow Spaghetti and Metro Maps to explain OCPM in a way that hopefully sticks in the reader's mind.

Object-Centric Process Mining

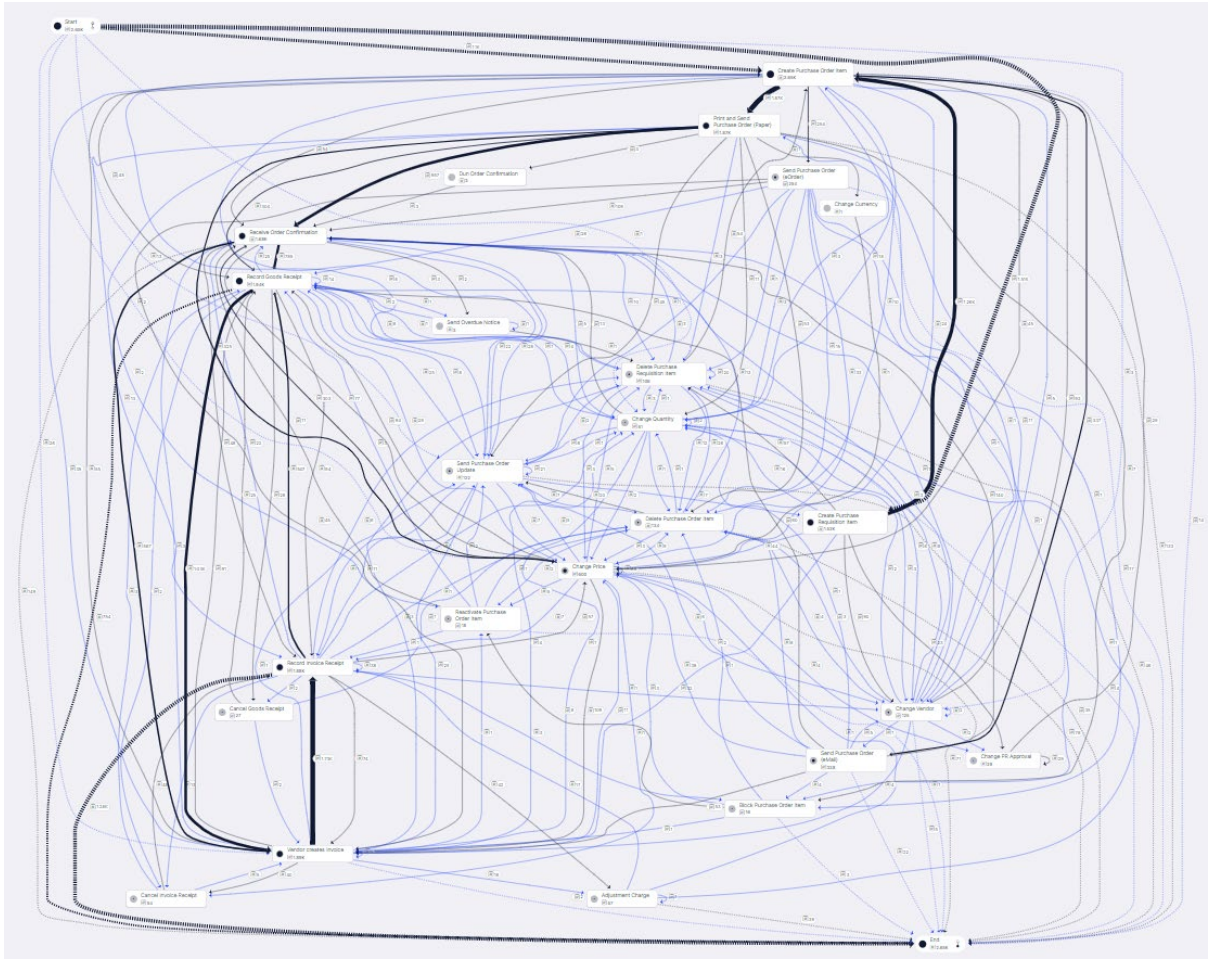
The goal of process mining is to analyze and improve processes using event data. Traditional process mining, while effective, has limitations such as time-consuming data extraction and transformation and difficulty in capturing interactions between objects like sales orders and invoices. Object-Centric Process Mining (OCPM) aims to overcome these challenges. Traditional process mining approaches assume that a process model describes the **lifecycle of a single object,** and each event refers to **precisely one object** (often called **case**) of a given type. These assumptions seem reasonable because most process-model notations do the same, and most end-users are familiar with these simplified notations. However, real-life activities may involve **multiple objects of different types**. Also, objects are not mutually independent and cannot be fully understood by looking at them in isolation. Objects are often related, e.g., an order has items and refers to a customer. As a result, extraction is time-consuming and needs to be repeated when new questions emerge. This is inflexible and prevents reuse. Also, logging is not system agnostic, i.e., the same business process creates different event data depending on the system used (e.g., SAP versus Oracle). Interactions between objects are not captured, and objects are analyzed in isolation. It is impossible to create views “on demand” when data are stored in traditional event logs. Fortunately, most of the existing process mining techniques can be lifted to multiple object types. The resulting advantages can be summarized as follows:

1. **Avoid repeatedly going back to your source systems** (a system-agnostic single source of truth).
2. **Avoid distortions due to the single-case assumption** (circumventing convergence and divergence problems).
3. **See and understand the interactions between different object types** (problems live at the intersections of processes and organizational entities).

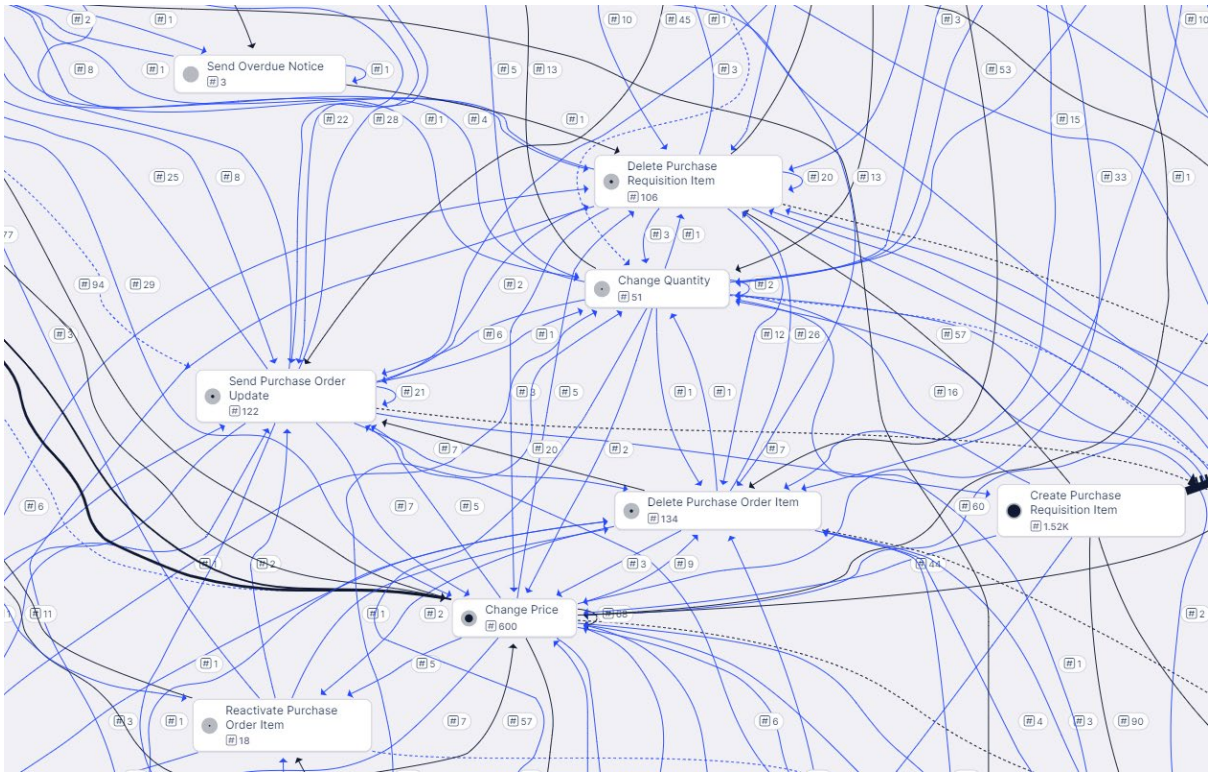
Anyone who has applied process mining at scale will immediately recognize the problems and see the need for OCPM. However, for people new to the topic, the above may all sound very abstract. Also, students and researchers who have learned about process mining, may not see the problem because they only use pre-cooked event logs prepared by others. The publicly available event logs are misleading because they do not show the problems extracting the data and the many design choices needed to do this. Therefore, I have repeatedly used the Rainbow Spaghetti and Metro Maps metaphors to explain the need for Object-Centric Process Mining (OCPM). This helped many to understand the concepts. Therefore, I wrote this article to help people who are new to the topic.

Rainbow Spaghetti

As mentioned, I have been using the **Spaghetti metaphor** to explain the complexity and variability of real-life processes. In the classical setting, each strand of Spaghetti (called "Spaghetto" in Italian) corresponds to one case, i.e., a sequence of activities related to one process instance (e.g., an order).



A Directly-Follows Graph (DFG) for a Purchase-to-Pay (P2P) process based on 2654 orders.



Part of the Directly-Follows Graph (DFG) for a Purchase-to-Pay (P2P) process showing more details.

The Spaghetti-like diagrams above illustrate the variability of real-life processes. However, this is only part of the story. Often, it is not recommended to limit the view to one object type (i.e., one case notion), because the root causes of problems often involve multiple objects of different types. While creating traditional case-based event logs, one faces the tradeoff between leaving out events and squeezing events into a single-case notion. For example, taking orders as a case notion, one needs to decide whether the scope includes the handling of individual order lines and deliveries or not. One order may consist of dozens of order lines, one delivery may refer to multiple order lines of different orders, and one order may have order lines that are part of multiple deliveries. This leads to the well-known c**onvergence and divergence problems**. Events referring to multiple objects of the selected type may get replicated, leading to unintentional duplication. The replication of events can lead to misleading diagnostics (convergence). There may also be multiple events that refer to the same case and activity, however, they differ with respect to one of the not-selected object types (divergence). In other words, events referring to different objects of a type not selected as the case notion become indistinguishable, looking only at the case and activity (i.e., event type). These problems show that one needs to **disentangle the Spaghetti** and consider the different types of objects explicitly. This leads to the notion of **Rainbow Spaghetti**. In Rainbow Spaghetti, we use **colors to distinguish the different object types**. Each Spaghetto (i.e., one strand of Spaghetti) is typed. This allows us to see the interactions between object types.

Image generated using DALL·E 3 showing the production process of a Porsche 911 using the Rainbow Spaghetti metaphor.

Image generated using DALL·E 3 showing a financial process using the Rainbow Spaghetti metaphor.

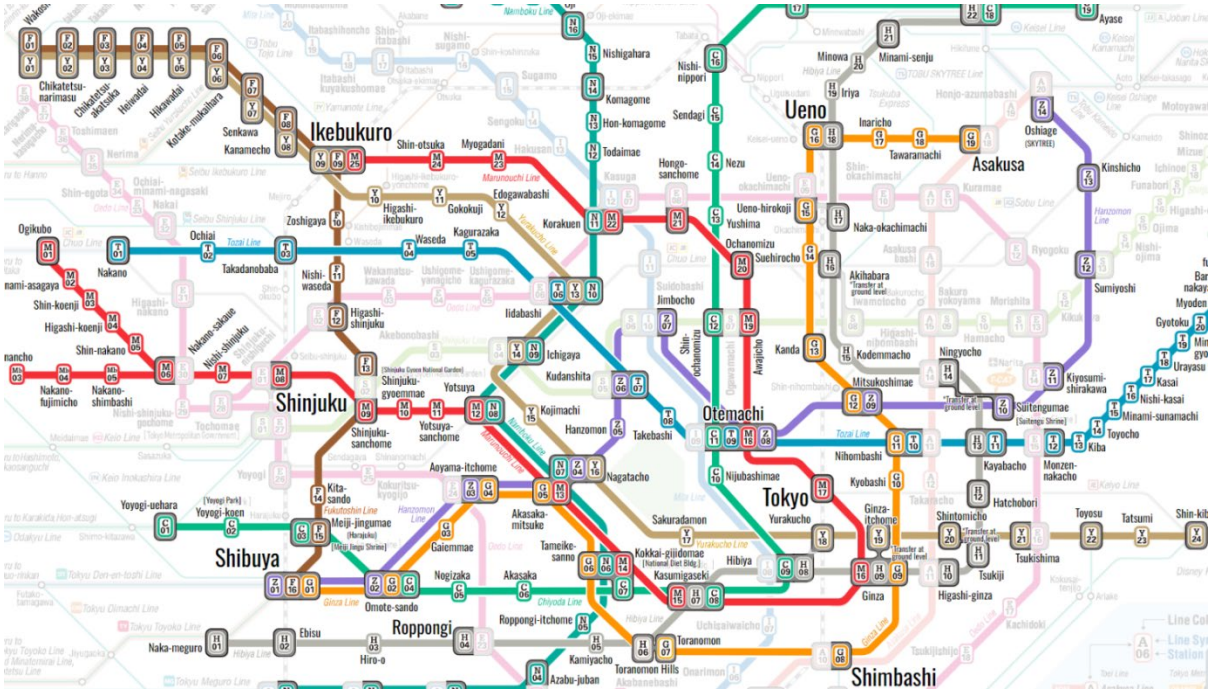
Image generated using DALL·E 3 showing a supply chain using the Rainbow Spaghetti metaphor.

The above images were generated using DALL·E 3 and illustrate the idea of Object-Centric Event Data (OCED). The colors refer to different object types. Examples of common object types include orders, items, packages, machines, employees, patients, customers, machines, containers, payments, containers, vehicles, rooms, etc. Note that the images show intertwined individual objects. When visualizing object-centric process models, we need to aggregate these object flows, e.g., arcs in

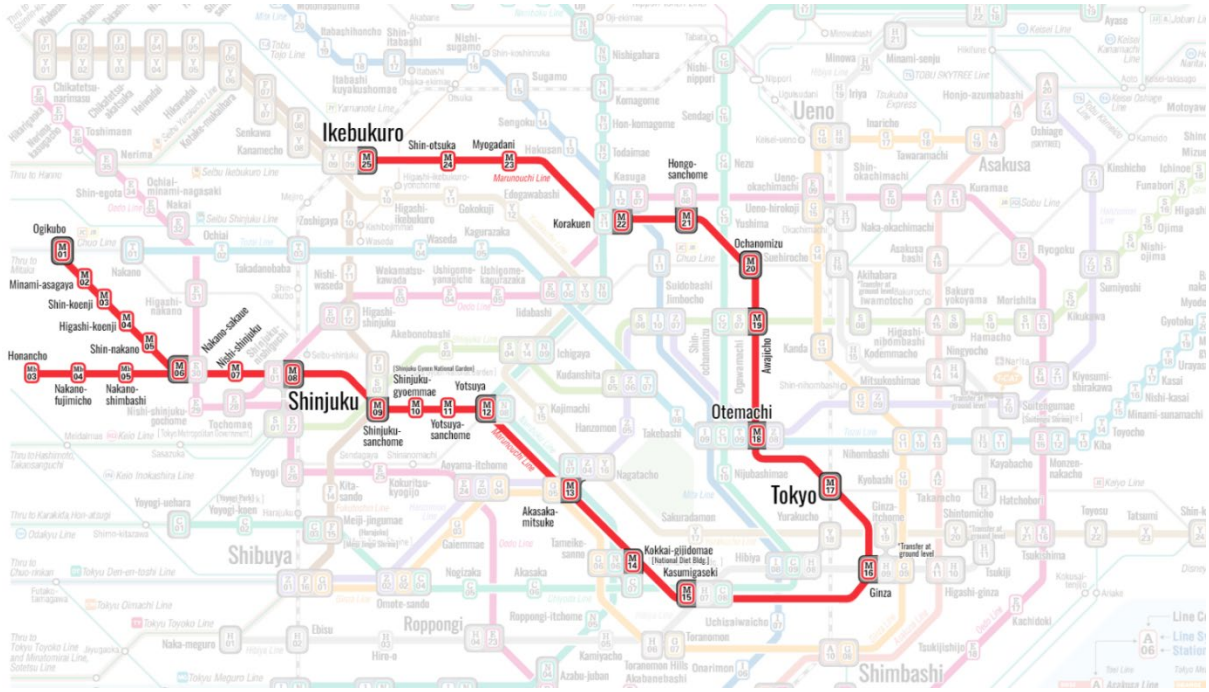
an **Object-Centric DFG**, places in an **Object-Centric Petri Net**, and gateways in an **Object-Centric BPMN** model.

Metro Maps

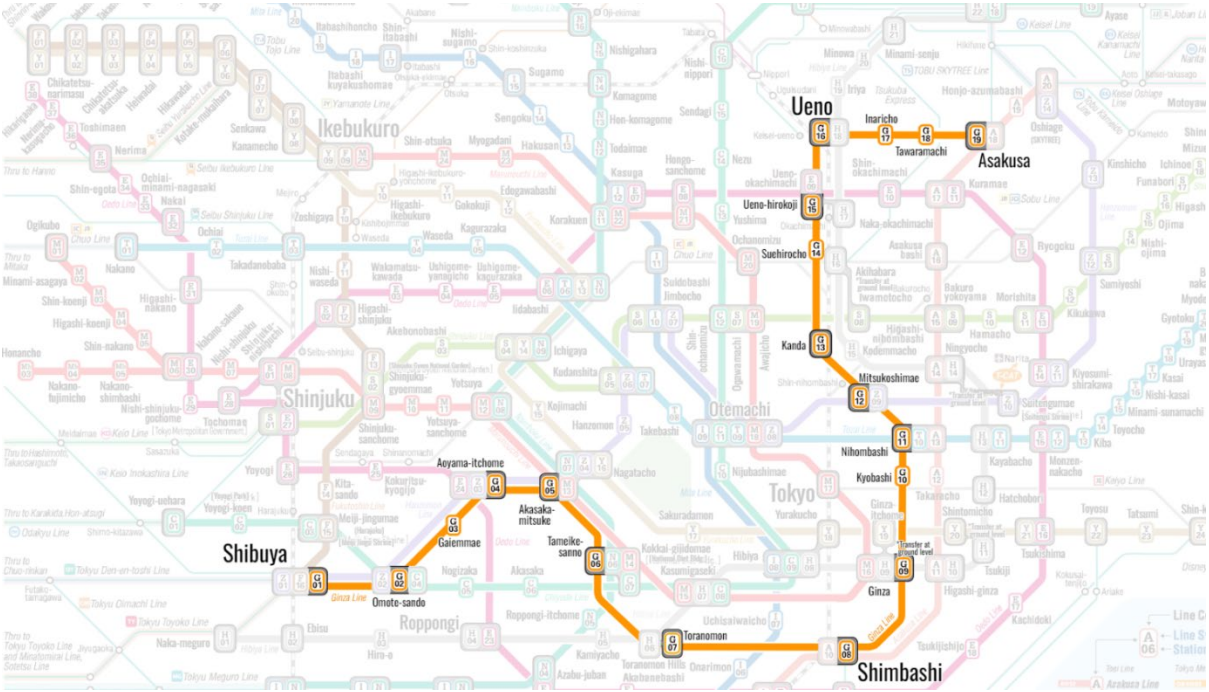
Another helpful metaphor to explain OCPM is a **Metro Map**. Whereas the Rainbow Spaghetti images generated DALL·E 3 were at the level of individual objects, a Metro Map is at **the level of object types**. Let's take a look at an example map.



A metro map of Tokyo showing nine metro lines. The metro lines correspond to object types, and the metro stations correspond to event types (i.e., activities). The map was created using the Tokyo Metro Co., Ltd. website.



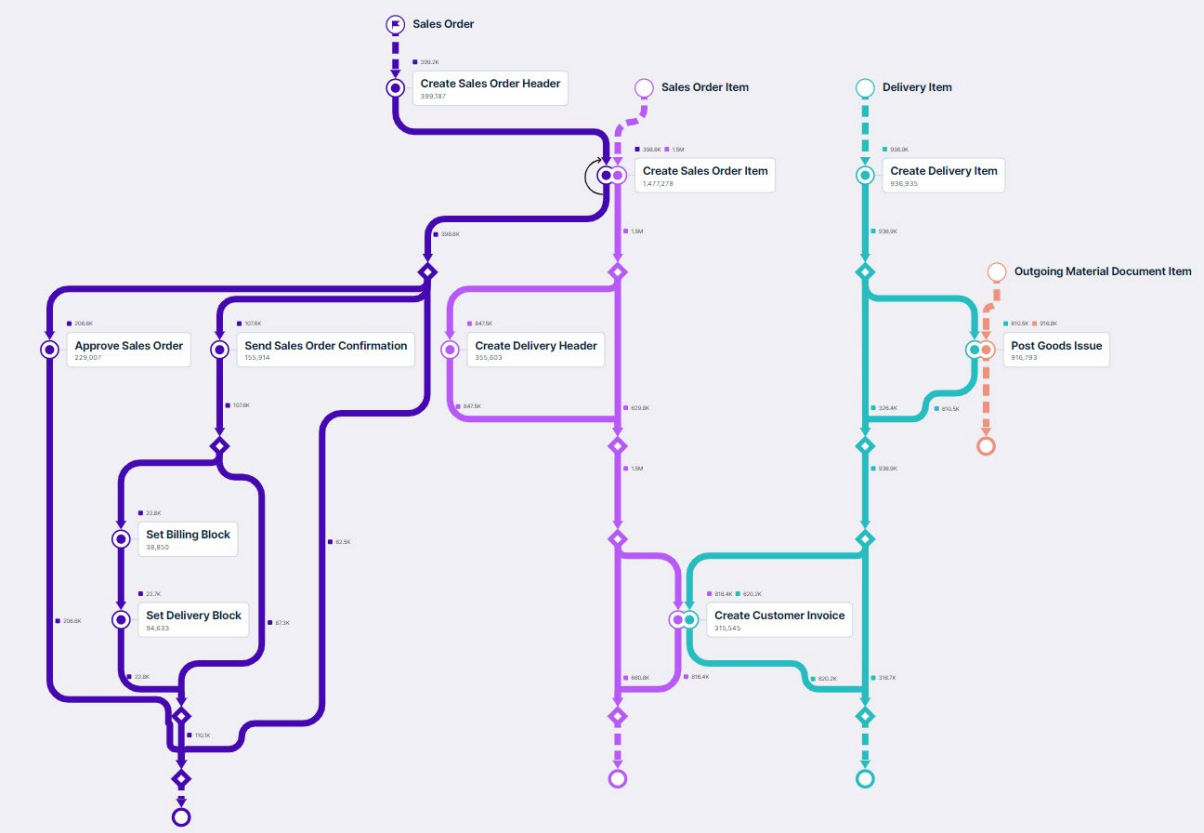
The Marunouchi line.



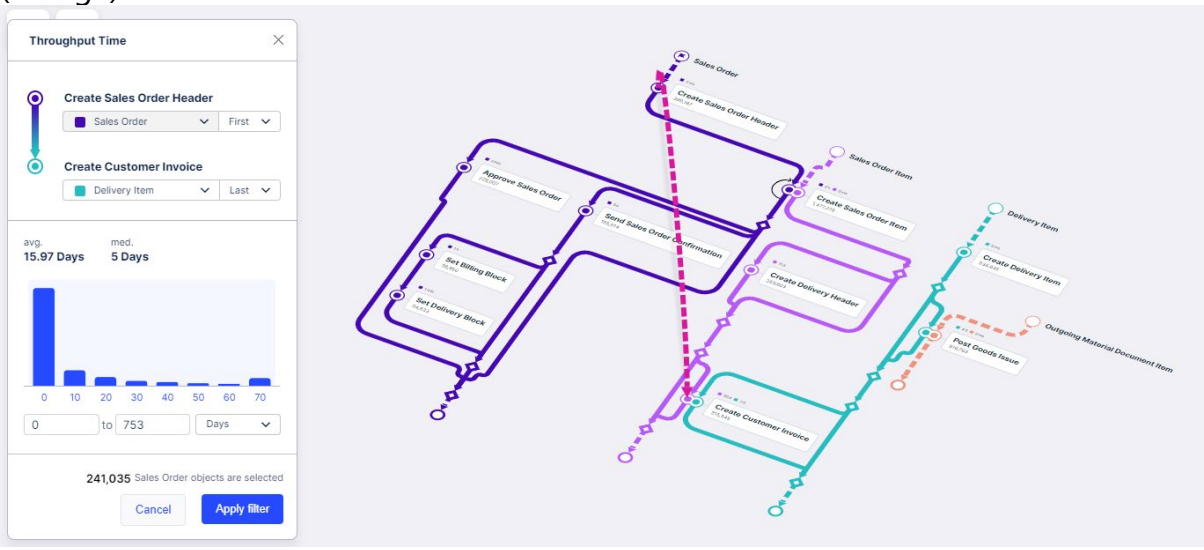
The Ginza line.

In the above maps, the **metro lines** correspond to **object types** and the **metro stations** can be seen as **activities**. Yuzuru Fukuda, the Corporate Executive Officer and Corporate Information Officer at Fujitsu, also used this metaphor at Celosphere 2023. This helped the audience to get the essence of OCPM.

Celonis provides a **Process Explorer** that discovers **Object-Centric DFGs** and **Process Adherence Manager** (called Process Sphere before) that discovers **Object-Centric BPMN** models and can be used for performance analysis.



Screenshot of the Process Adherence Manager (called Process Sphere before) showing a discovered Object-Centric BPMN model with four object types: sales orders (blue), sales order items (purple), delivery items (green), and outgoing items (orange).



Screenshot of the Process Adherence Manager showing performance diagnostics.

The screenshots of the **Celonis Process Adherence Manager** show a striking similarity with Tokyo's **Metro Maps** shown before. When computing the time to go from address A to address B, one often needs to consider **multiple** metro lines. Similarly, when analyzing process performance, one also needs to consider multiple object types.

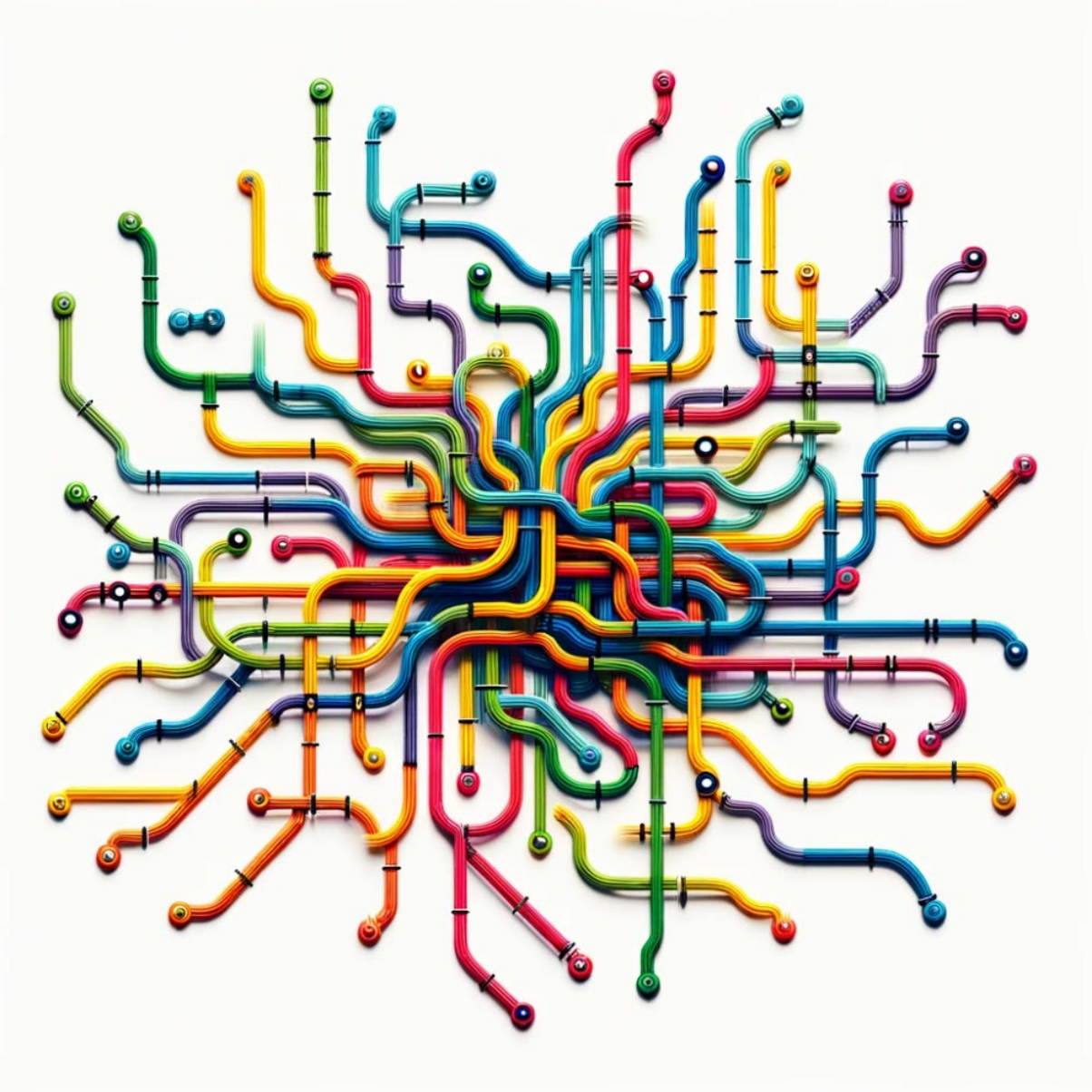


Image generated using DALL·E 3 combining the Rainbow Spaghetti and Metro Map metaphors.

I have often used the **Google Maps** metaphor to explain process models overlayed with performance and compliance information. Bottlenecks in operational processes correspond to traffic jams. This immediately makes the point that enriching process models with real-time information about the processes is vital. This way, we can create a living **Digital Shadow** of an organization. However, the Google Maps metaphor works best for traditional process-mining techniques using the single-case notion. The metro maps make the additional point that different object types (i.e., metro lines) are interacting. Problems with one metro line may cause problems in other parts of the network.

What Has Been Seen Cannot Be Unseen!

Metaphors can be used to express complex concepts, such as Object-Centric Process Mining (OCPM), by exaggerating critical characteristics in a visual manner. In this article, I used Rainbow Spaghetti to show the need for object-centric event data and Metro Maps to convey the idea of object-centric process models. Hopefully, this helps to stick the idea of OCPM in the reader's mind. Because traditional process modeling notations use a single case notion, many practitioners have developed a **blind spot, unable to see the true fabric of operational processes**. In an organization, often multiple source systems are used (e.g., software systems from SAP, Salesforce, Microsoft, and Oracle), and one system may store data scattered over thousands of database tables. Tables in these systems are connected but seldom in a one-to-one manner. Just like the real "process fabrik", also the "data fabrik" is object-centric. OCPM helps organizations create a single source of truth that is system-agnostic. It is possible to dynamically create views by selecting object types and activities without returning repeatedly to the source systems. This also helps to avoid the usual distortions caused by squeezing data into cases. By truly understanding the interaction between objects, one can see and address the real root causes of performance and compliance problems. Hopefully, the adage "What has been seen cannot be unseen" holds after reading this article.

References

1. W.M.P. van der Aalst. *Process Mining: Data Science in Action*. Springer-Verlag, Berlin, 2016.
2. W.M.P van der Aalst. *Object-Centric Process Mining: Unraveling the Fabric of Real Processes*. Mathematics, 11(12):2691, 2023.
3. W.M.P van der Aalst, *Object-Centric Process Mining: The Next Frontier in Business Performance*, 2023. celon.is/OCPM-Whitepaper.
4. W.M.P van der Aalst, *TomTom for Business Process Management (TomTom4BPM)*. Advanced Information Systems Engineering. CAiSE 2009. Lecture Notes in Computer Science, vol 5565. Springer, Berlin, Heidelberg.
5. W.M.P. van der Aalst and A. Berti. *Discovering Object-Centric Petri Nets*. Fundamenta Informaticae, 175(1-4):1–40, 2020.
6. W.M.P. van der Aalst and J. Carmona, editors. *Process Mining Handbook*, volume 448 of Lecture Notes in Business Information Processing. Springer-Verlag, Berlin, 2022.
7. G. Acampora, A. Vitiello, B. Di Stefano, W. van der Aalst, C. Günther, and E. Verbeek. *IEEE 1849: The XES Standard - The Second IEEE Standard* Sponsored by IEEE Computational Intelligence Society. IEEE Computational Intelligence Magazine, 12(2):4–8, 2017.
8. A. Berti, I. Koren, N. Adams, G. Park, B. Knopp, N. Graves, M. Raffei, L. Liss, L. Tacke Genannt Unterberg, Y. Zhang, C. Schwanen, M. Pegoraro, W.M.P. van der Aalst, *OCEL (Object-Centric Event Log) 2.0 Specification*. https://www.ocel-standard.org/2.0/ocel20_specification.pdf (2023)

9. A.F. Ghahfarokhi, G. Park, A. Berti, and W.M.P. van der Aalst. *OCEL Standard*. www.ocel-standard.org, 2020.
10. IEEE Task Force on Process Mining: *Process Mining Manifesto*. https://www.tf-pm.org/resources/manifesto (2011)
11. Process and Data Science Group: *Object-Centric Event Log (OCEL 2.0)* Website Providing Data Sets, Software, Etc. www.ocel-standard.org (2023)

About the author

Prof.dr.ir. Wil van der Aalst is a full professor at RWTH Aachen University, leading the Process and Data Science (PADS) group. He is also the Chief Scientist at Celonis, part-time affiliated with the Fraunhofer FIT, and a member of the Board of Governors of Tilburg University. His research interests include process mining, Petri nets, business process management, workflow management, process modeling, and process analysis. Wil van der Aalst has published over 900 articles and books. According to Research.com, he is the highest-ranked computer scientist in Germany and ranked 10th worldwide. According to Google Scholar, he has an H-index of 177 and more than 140.000 citations. Van der Aalst is an IFIP Fellow, IEEE Fellow, ACM Fellow, and received honorary degrees from the Moscow Higher School of Economics (Prof. h.c.), Tsinghua University, and Hasselt University (Dr. h.c.). He is also an elected member of the Royal Netherlands Academy of Arts and Sciences, the Royal Holland Society of Sciences and Humanities, the Academy of Europe, the North Rhine-Westphalian Academy of Sciences, Humanities and the Arts, and the German Academy of Science and Engineering. In 2018, he was awarded an Alexander-von-Humboldt Professorship.

Originally published as a LinkedIn article https://www.linkedin.com/pulse/rainbow-spaghetti-metro-maps-object-centric-process-wil-van-der-aalst-ebn2e 26-11-2023